149 447 169 507

449 267 267 267 109 327 327 387

SCA SCA TTC V GTG CGA CCC CCC CCC CCC CCC 266 266 276 276 276 260 660 676 676 676 676 676 GAAA CGC CGC T ACA ACA ACA CGC GCC GCC S CGC CGC CGC CTG CTG CTG CTG CTG CTG CGC CAC CAC CAC CAC CAC CAC CAA CCAC AAAC AAAC GCC CTC T AACT ACT CTC CTC CAA CAA CTG CTC CTC CTC CGC CGC CGC CGC GAG GAG GAG CTG V GTG GTG CTG CTG CTG CTG CTG CTG CTG A A A G CTG CGC CAC CTG CTG CTG CTG CGG CGG GCG GCG GCA GCA I I ATC CAC CAC CCGG GAC GTC GTC CGC CGC GCT GCT GGG CGG 

FIG. 1A.

209 627 627 687 687 747 747 807 807 909

V GTC K AAA E GAG CTC I I ATC V TGG Y TAC Y TTC TTC V OTG S S S AGT AGC GCC CTC CTC CTC ATC ATC CTG CTG CTC CTC E GAG GTG S AGC CGC CGC GGA I ATT GTA GTA TTC TTC P SCT

V
37C

ACC

CTC

CTC

CTC

CTC FROM FIG.

JG CCC ACC

TG CCA GGA

JG AAG AGG

JG CCC ATC

TO TG CCG

TO TG CCG

TO TG CCG W TGG V GTG AGG CAG CAG 1 ATT S S TCA CAG CAG CAG V CAG CTC CTG
FTC
A
GCA
S
TCC
A
A
CCA
A
CTC
CTG ACA N N AAC AAG GTG GTG GAC I C ACT TGC ACT ATC ACC ACCA ATC CGC ACCA ATC CGC ACCA ACC CACA ACC E I S
GAA ATT TCG A
S F TT GTT A
ATT TTA CAG A
S H Q
AGC CAC CAG A
ATG GTC TCC
Q N F
CAG AAC TTC

FROM FIG. 1B

F A N S A L N P' I L Y N M T L C RG AGG AAT GAG 987

K K I F C C F W F P E K G A I L T D T 349

AAG AAA ATT TTT TGC TGC TTC TGG TTC CCA GAA AAG GGA GCC ATT TTA ACA GAC ACA 1047

V K R N D L S I I S G 386

GTC AAA AGA AAT GAC TTG TCG ATT ATT TCT GGC TAA 1088

TCCATCTTAAAAAAAAAAAAAAAAAGATTTGTTATGGGTTCCTTTTAAATGTGAACTTTTTAGTGTGTTTTGTAATATG CTCTTGAACCTGGGAGGCAGAGGTTGCAGTGAGCCGAGATCGTGCCATTGCACTCCAACCAGGGCAACAAGAGTGAAAC GTTTCATAATATTTTCCCTTTATAAAAGGATTTGTTGGCCAGGTGCAGTGGTTCATGCCTGTAATCCCAGCAGTTTGGG AGGCTGAGGTGGGTTGATL ACCTGAGGTCAGGAGTTCGAGACCAACCTGACCAACATGGTGAGACCCCGTCTCTACTA TTTTCTTTATAGCCGAGTTTCTCACACCTGGCGAGCTGTGGCATGCTTTTAAACAGAGTTCATTTCCAGTACCCTCCA TCAGTGCACCCTGCTTTAAGAAAATGAACCTATGCAAATAGACATCCACAGCGTCGGTAAATTAAGGGGGTGATCACCAA AAAATAAAAAAAAAATTAGCTGGGAGTGGTGGTGGCACCTGTAATCCTAGCTACTTGGGAGGCTGAACCAGGAGAAA 

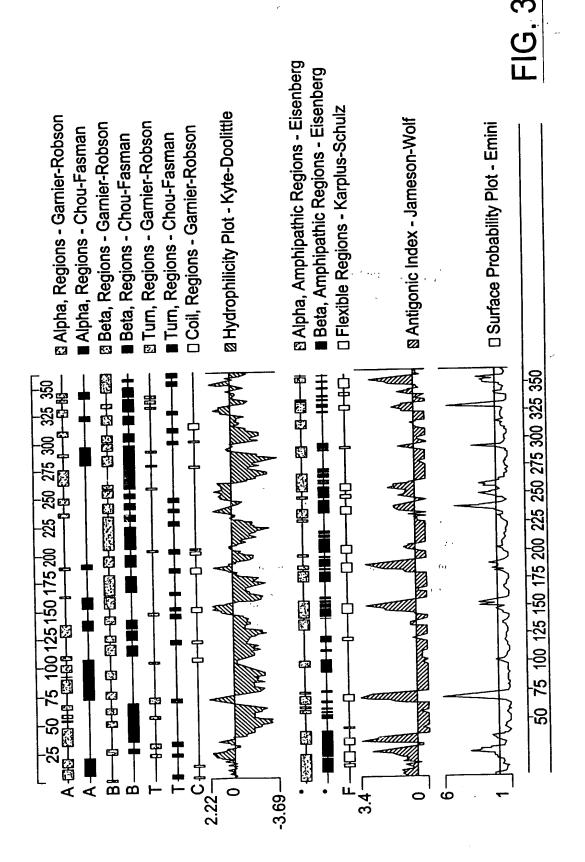
RTA01/2057957v1

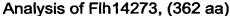
=1G. 1C.

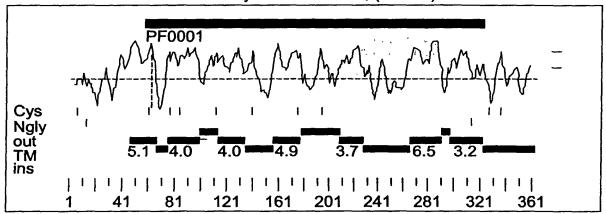
### 4/15

```
Sequence Description
                                                         E-value
7tm_1
           PF00001 7 transmembrane receptor (rhodopsin 119.9
                                                         4. 7e-37
Parsed for domains:
Sequence Domain seq-f seq-t hmm-f hmm-t
                                             score
                                     259 [] 119.9
                 57 321
7tm 1
          1/1
Alignments of top-scoring domains:
7tm_1: domain 1 of 1, from 57 to 321: score 119.9, E = 4.7e-37
                *->GN1LVi1vi1rtkk1rtptnifi1NLAVADLLf11t1ppwa1yy1vg
                   GN+ ++++++ +++ ++ ++ ++ ++ ++ ADLLf + p++ ++ -+
                   GNVCALVLVAR-RRRGATACLVLNLFCADLLFISAIPLVLAVR-WT 101
  F1h14273, 57
                gaadWpfGsa1Ck1vtaldvvnmyaSi11Lta1SiDRY1A1vhP1ryrrr
                   e W++G++ C+1+ ++++++ + 11+L+a S++R + Iv 1+ +r
  F1h14373, 102 -- EAWLLGPVACHLLFYYMTLSGSVTILTLAAVSLERMVCIV-HLQRGVR 148
                rtsprrAkvvillvwvlallls1Pp11fswvktveegngt1nvnvevCli
                   fr +v+++1+W +++++1P +f+ v+ ++ ++ ++ +C++
  F1h14273, 149 GPGRRARAVLLALIVGYSAVAALPLCVFFRVVPQRLPG--ADQEISICTL 196
                dfpccstasvstwlrsyv11st1wgF11P11vilvcYtrI1rt1r.....
                             ++5+ +++ ++ F1+P 1vi++ Y+ I1 + + ++++
  F1h14273, 197 IWPTIPG-----EISWDVSFVTLNFLVPGLVIVISYSKILQITKasrkr 240
                   .....kaakt11vvvvvFv1CW1Pyfiv111dt1c
                 + + +++++ + ++++ ++ +t1++++v F++ W P +++11 +
  F1h14273, 241 icvslayseehgirvsqqdfRLFRTLFLLMVSFFIMWSPIIITILLILIQ 290
                 .1siimsstCelerv1pta11vt1wLayvNsc1NPi1Y<-*
                          + + p ++++ + ++++Na+1NPi+Y
  F1h14273, 291 nFK-----QDLVIWPSLFFWVVAPTFANSALNPILY 321
```

FIG. 2.







>F1h14273, 1086 bases, 1825 checksum.
MSPECARAAQDAPLRSLEQANRTRFPFFSDVKGDHRLVLAAVETTVLVLEFAVSLLGNVC
ALVLVARRRRRGATAQLVLNLFCADLLFYSAIPLVLAVRWTEAVLLGPVACHLLFYVMTL
SGSVTILTLAAVSLERMVCIVHLQRCVRGPGRRARAVLLALIWGYSAVAALPLCVFFRVV
PQRLPGADQEISICTLIWPTIPGEISWDVSFVTLNFLVPGLVIVISYSKILQITKASRKR
LTVSLAYSESHQIRVSQQDFRLFRTLFLLMVSFFIMWSPIIITILLILIQNFKQDLVIWP
SLFFWVVAFTFANSALNPILYNMTLCRNEWKKIFCCFWFGPEKGAILTDTSVKRNDLSIIS
G+

FIG. 4.

Prosite Pattern Matches for F1h14273

>PS00001/PDUC00001/ASN\_GLYCUSYLATIUN N-glycosylation site.

Query: 21

NRTR 24 NMTL 325 Query: 322

>PS0004/PDDC00004/CAMP\_PHDSPHD\_SITE cAMP- and cGMP-dependent protein kinase phosphorylatic

KRLT 242 Query: 239

><u>PS00005</u>/PDEC00005/PKC\_PHESPHE\_SITE Protein kinase C phosphorylation site.

SRK SVK Query: 237

><u>PS00006</u>/PDDC00006/CK2\_PHDSPHD\_SITE Casien kinase II phosphorylation site.

SQQD 259 Query: 256 >PS00008/PDDC00008/MYRISTYL N-myristoylation site.

GNVCAL 62 GATACL 77 GAILTD 348 Query: 57 Query: 72

Query: 343

>PS00009/PDDC00009/AMIDATION Amidation site.

PCRR 153 Query: 150 >PS00029/PDDC00029/LEUCINE\_ZIPPER Leucine zipper pattern. 127 LGPVACHLLFYVMTLSGSVTIL Query: 106

8/15

# Transmembrane Segments Predicted by MEMSAT

Start	End	Orient	Score	
46	66	out->ins	5.1	
75	98	ins->out	4.0	
113	134	out->ins	4.0	
156	177	ins->out	4.9	
209	227	out->ins	3.7	
266	289	ins->out	6.5	
297	321	out->ins	3.2	

# >F1h14273,

MSPECARAAGDAPLRSLEQANRTRFPFFSDVKGDHRLVLAAVETTVLVLIPAVSLLGNVC ALVLVARRRRRGATACLVLNLFCADLLFISAWPLVLAVRWTEAWLLGPVACHLLFYVMTL SGSVTILTLAAVSLERMVCIVHLQRGVRGPGRRARAVLLALIWGYSAVAALPLCVFFRVV PQRLPGADQEISICTLIWPTIPGEISWDVSFVTLNFLVPGLVIVISYSKILQITKASRKR LTVSLAYSESHQIRVSQQDFRLFRTLFLLMVSFFIMWSPIIITILLILIQNFKQDLVIWP SLFFWVVAFTFANSALNPILYNMTLCRNEWKKIFCCFWFPEKGAILTDTSVKRNDLSIIS G

## Transmembrane Segments for Presumed Mature Peptide

Start	End	Orient	Score	
14	37	ins->out	4.0	
52	73	out->ins	4.0	
95	116	ins->out	4.9	
148	166	out->ins	3.7	
205	228	ins->out	6.5	
236	260	out->ins	3.2	

## >F1h14273, \_mature

LVLVARRRRRGATACLVLNLFCADLLFISAIPLVLAVRWTEAWLLGPVACHLLPYVMTLS GSVTILTLAAVSLERMVCIVHLQRGVRGPGRRARAVLLALIWGYSAVAALPLCVFFRVVP QRLPGADQEISICTLIWPTIPGEISWDVSFVTLNFLVPGLVIVISYSKILQITKASRKRL TVSLAYSESHQIRVSQQDFRLFRTLFLLMVSFFIMWSPIIITILLILIQNFKQDLVIWPS LFFWVVAFTFANSALNPILYNMTLCRNEWKKIPCCFWFPEKGAILTDTSVKRNDLSIISG Input file 14273m; Output File 14273mtrα Sequence length 1560

TTGCCAAGCTCAGCGTAAGCCTCTTCCACTGCAATCTCACAGAAGGGGTTCATGGAGTGCTTCACACCATCAGTGACCA

130 390 150 450 50 150 210 210 270 270 330 GAT I ATC R CGC S AGC V GTC ACG S TCG L CTC CGG TC CGG GTC CTG F TTC GGA R CGT F TT T ACA AGC F L CTG CGC CGC CTC CTC L L L CTC P CCT

V GIT

A GCG

L

TTG

TTG

S

GCC

S

S

TCC

TCC F TTC ACC V GTG GAT L CTG ACG ACG CTG TAC H CAC T L CTG GCG GCG GCG GCG GCG S M SGCCATCTTCCCGGACGCGTGGGCCGGGCGCCCGGC ATG CAA LL TTG V GTG CTC CTC CTC CTC CTG CTG GGC GTC V GTG GTG ACC CGG CTG S AGC CTC TAC TAC CGC CAC CAC CLG CTG GCC V GTG GTG S TCG JBC SAC TCA TCA CTA CTC GCT K AAG V GTC GGG GGG GGG GGG CAC CAC 26T 37C 37C 37C 36T 56T 56G 66G

190	J	210 630	230 690	250	270 810	290 870	310 930	330	350 1050	362 1086
ш <del>{</del>		F TTT	I ATT	S AGC	M ATC	CAA	T ACG	<b>₩</b> TGG	S TCT	
ر د ک	CAG	۷ 576	AAA AAA	E GAG	L CTC	$_{\rm ATC}^{\rm I}$	F 77C	E GAA	T ACG	
	S P C	D GAT	S 700	S TCT	L CTG	L TTG	A	N AAC	D GAT	
	_	•	•	≺ TAC						
ۍ ر	טטר	STCA	S AGT	GCA	L CTC	L CTC	V GTG	FTC	7 11	
•	_	_		L TTG						
				S AGC						
				L CTG						
Q (	CAG	I ATA	^ 10	T ACG	L CTC	I ATC	٦ 1	AAC	A AAG	TAA
٥.		80 00 00		L CTT						S AGC
>		AAC	_	R AGG						S TCC
>	010	۳ 23	P CCG	AAG	D GAC	٩ 2 3	<b>≯</b> 00 100	$\frac{1}{I}$ ATA	FTT	I ATT
œ				CGG						\ GTT
لبا	110	D GAT		S 70G	CAA CAA	¥ 766	> GTC	AAC	F TTC	S TCT
										L
1	_	_	_							D GAC
	•	•		ACG						
	_									R CGA
										AGG
_	CTG	I	- 4-F	T	CAC	> \ 	A P C	F T T T	R AGG	۷ 570

7A.

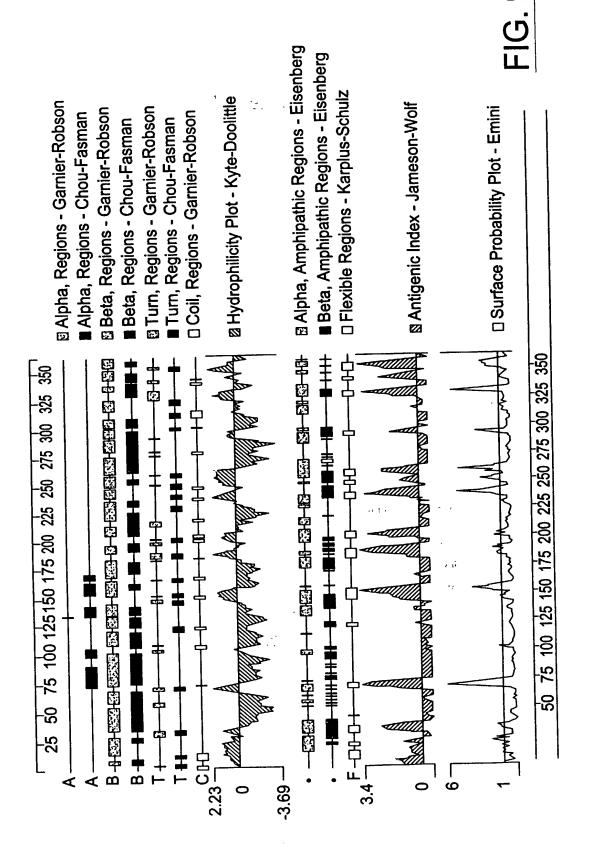
FROM FIG.

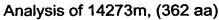
CTAGCCTCTGGTGCCAGGTGAACCACGGTGTGCATGTAAAGCGAGTTAACTTCAAGGAAAGCCCACCAGTGCGCCCTGC TTTAAAAATACCCGACTTCCAACAGCAGGCATCTACGGAGCCAGCAAATTAAGGAATGATCGCTCAGTATAAAAATAT TTTCCTTAAAAGAACTTTCTATGGGTTCCTTTTGTGAACTTTTTTAAGTGTGTTTGTAATATGATCTAGTTAAAATI 

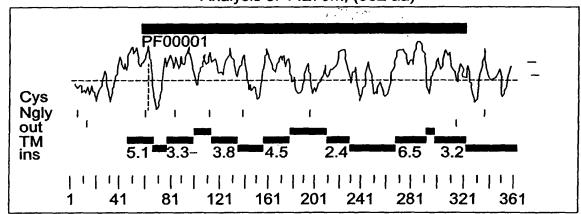
### 11/15

```
Query:
             14273m,
Scores for sequence family classification (score includes all domains):
Sequence Description
                                                  Score
           PF00001 7 transmembrane receptor (rhodopsin 118.8
                                                         1e-35
7tm 1
Parsed for domains:
Sequence Domain seq-f seq-t hmm-f hmm-t
                                             score
                       321 ..
                                     259 [ ] 118.8
                                                    1e-36
                  57
                                1
Alignments of top-scoring domains:
7tm_1: domain 1 of 1, from 57 to 321: score 118.8, E = 1e-36
                *->GN1LVilvilrtkk1rtptnifilNLAvADLLf11t1ppwa1yy1vg
                   GN+ ++++++ ++++ ++ +++1NL ADLLf+ + D++ ++ ++
                   GNVCALVLVAR-RRRRGASASLVLNLFCADLLFTSAIPLVLVVR-WT 101
   14273m, 57
                gaadWpfGaalCklvtaldvvnmyaSilllLtalSiDRYlAlvhPlryrrr
                   e W++G+++C+1+ ++++++ + i1+L+a S++R + Iv 1r
   14273m, 102 -- EAWLLGPVVCHLLPYVMTMSGSVTILTLAAVSLERMVCIV-RLRRGLS 148
                rtsprrA, kvvi11vWv1a111s1Pp11fswvktveagngt1nvnvtvC1
                   rr++++++++\\ ++1++1P ++++ v +
   14273m, 149 GP-GRRTqAALLAFIWGYSALAALPLYILFRVVPQRLPGGD--QEITPCT 195
                 idfpeestasvstw1rsyv11st1vgF21P11vi1voYtr11rt1r....
                              ++5+ +++ ++ F1+P 1vi++ Y+ 11 + + +++
                 +d+p++ +
   14273m, 196 LDWPNRIG-----EISWDVFFETLNFLVPGLVIVISYSKILQITKasrk 239
                   .....kaakt11vvvvvFv1CW1Pyfiv111dt-
                 + + + +++++ + ++++ ++ +t1++++v F++ W P 1++11 +
  14273m, 240 ritislayseshqirvsqqdyRLFRTLFLLMVSFFIMWSPIIITILLILI 289
                 c. lalimestCelerv1ptallvt1wLayvNsc1NPi1Y<-*
                              + p ++++ + ++++Ns+1NPi+Y
  14273m, 290 QnFR-----QDLVIWPSLFFWVVAFTFANSALNPILY 321
```

FIG. 8.







>14273m, 1086 bases, 6943 checksum.
MSPECAQTTGPCPSHTLDQVNRTHFPFFSDVKGDHRLVLSVVETTVLGLIFVVSLLGNVC
ALVLVARRRRRGASASLVLNLFCADLLFTSAIPLVLVVRWTEAWLLGPVVCHLLFVVMTM
SGSVTILTLAAVSLERMVCIVRLRRGLSGPGRRTQAALLAFIWGYSALAALPLYILFRVV
PQRLPGGDQEIPICTLDWPNRIGEISWDVFFETLNFLVPGLVIVISYSKILQITKASRKR
LTLSLAYSESHQIRVSQQDYRLFRTLFLLMVSFFIMWSPIIITILLILIQNFRQDLVIWP
SLFFWVVAFTFANSALNPILYNMSLFRNEWRKIFCCFPFPEKGAIFTDTSVRRNDLSVIS
S\*

FIG. 10.

#### 14/15

Prosite Pattern Matches for 14273m,

>PS00001/PD0C00001; ASN\_GLYCOSYLATION N-glycosylation site.

Query: 21

NRTH

24

Query: 322

NRTH

325

242

>PS00002/PD0000002/GLYCOSAMINOGLYCAN Glycosaminoglycan attachment site.

Additional rules:

ŔŬ There must be at least two acidic amino acids (Glu or Asp) from -2 to -4 relative to the serine.

Query: 148

SGPG 151

>PS00004/PD0C00004/CAMP\_PHOSPHO\_SITE cAMP- and cGMP-dependent protein kinase phosphorylation

Query: 239

KRLT

>PS00005/PD0C00005/PKC\_PH0SPH0\_SITE Protein kinase C phosphorylation site.

Query: 237

SRK 239

SVR Query: 350 352

>PS00006/PD0C00006/CK2\_PH0SPH0\_SITE Casein kinase II phosphorylation site.

Query: 40

SVVE

43

Query: 256

SQQD 259

>PS00008/PDDC00008/MYRISTYL N-myristoylation site.

Query: 57

GNVCAL 62

Query: 72

GASASL 77

Query: 343

GAIFTD 348

>PS00009/PDDC00009/AMIDATION Amidation site.

Query: 150

**PGRR** 

153

FIG. 11.

15/15

## Transmembrane Segments Predicted by MEMSAT

Start	End	Orient	Score	
46	66	out->ins	5.1	
77	98	ins->out	3.3	
113	134	out->ins	3.8	
156	177	ins->out	4.5	
209	227	out->ins	2.4	
266	289	ins->out	6.5	
297	321	out->ins	3.2	

# >14273m,

MSPECAQTTGPGPSHTLDQVNRTHFPFFSDVKGDHRLVLSVVETTVLGLIFVVSLLGNVC ALVLVARRRRGASASLVLNLFCADLLFTSAIPLVLVVRWTEAWLLGPVVCHLLFYVMTM SGSVTILTLAAVSLERMVCIVRLRRGLSGPGRRTQAALLAFIWGYSALAALPLYILFRVV PQRLPGGDQEIPICTLDWPNRIGEISWDVFFETLNFLVPGLVIVISYSKILQITKASRKR LTLSLAYSESHQIRVSQQDYRLFRTLFLLMVSFFIMWSPIIITILLILIQNFRQDLVIWP SLFFWVVAFTFANSALNPILYMMSLFRNEWRKIFCCFFFPEKGAIFTDTSVRRNDLSVIS

### Transmembrane Segments for Presumed Mature Peptide

Start	End	Orient	Score	
16	37	ins->out	3.3	
52	73	out->ins	3.8	
95	116	ins->out	4.5	
148	166	out->ins	2.4	
205	228	ins->out	6.5	
236	260	out->ins	3.2	

### >14273m, \_mature

LVLVARRRRRGASASLVLNLFCADLLFTSAIPLVLVVRWTEAWLLGPVVCHLLFYVMGMS GSVTILTLAAVSLERMVCIVRLRRGLSGPGRRTQAAIIAFIWGYSALAALPLYILFRVVP QRLPGGDQEIPICTLDWPNRIGEISWCVFFETLNFLVPGLVIVISYSKILQITKASRKRL TLSLAYSESKQIRVSQQDYRLFRTLFLLMVSFFTMWSPIIITILLILIQNFRQDLVIWPS LFFWVVAFTFANSALNPILYNMSLFRNEWRKIFCCFFFPEKGAIFTDTSVRRNDLSVISS